DOCUMENT RESUME

ED 190 593 TM 800 379

TITLE	Alternative Conceptions of Competence Assessment. Proceedings of the 1979 Annual CAPT Conference.
INSTITUTION	Northwest Regional Education Lab., Portland, Oreg. Clearinghouse for Applied Performance Testing.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Nov 79
NOTE	63p.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; *Competency Based Education; Evaluation Criteria; *Evaluation Methods; *Experiential Learning; *Informal Assessment; Medical Education; Nontraditional Students; Outcomes of Education; Secondary Education; *Simulation; *Student Evaluation

ABSTRACT

Several new developments in educational measurement which contribute to a broadening of assessment options were presented at the conference on Alternative Conceptions of Competence. A common theme synthesizing these alternatives was a view of the complexity of human competence and the inadequacy of the multiple-choice examination as a measure of competence. The conference was divided into two parts: the first focusing on the conceptual basis for alternative competence measures and the second dealing with application of nontraditional assessment techniques in particular contexts for particular purposes. The first speaker, David G. Winter, proposed a set of four characteristics that any competency measure ought to have. Christine H. McGuire examined the use of simulation techniques in medical education to assess four critical components of competencies: observational/interpretive skills, problem solving skills, interpersonal/communication skills, and technical skills. The third speaker, Georgine Loacker, described the way Alverno College redefined the nature and function of student assessment in a liberal arts Curriculum. The last speaker, Thomas R. Owens, reviewed the difficulties inherent in measuring the outcomes of experiential learning programs at the secondary level. (Author/GSK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190593

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Alternative Conceptions of Competence Assessment

Proceedings of the 1979 Annual CAPT Conference

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
M. Thorne

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800379

CAPT
2

ALTERNATIVE CONCEPTIONS OF COMPETENCE ASSESSMENT

Proceedings of the 1979 Annual CAPT Conference

Prepared by:

The Clearinghouse for Applied Performance Testing
Assessment & Measurement Program
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204

November 1979

JUN 2 0 1980

This work is published by the Clearinghouse for Applied Performance Testing (CAPT) of the Northwest Regional Educational Laboratory, a private nonprofit corporation. The work contained herein has been developed under a grant with the National Institute of Education (NIE), Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position of the National Institute of Education, and no official endorsement by the Institute should be inferred.

CAPT Director: Beverly Anderson

NIE Project Monitor: Judy Shoemaker

Advisory Board:

Richard Brickley
Thomas Corcoran
Connie Kravas
Christine McGuire

TABLE OF CONTENTS

Page

Preface i

New Approaches to Competence Assessment 1
David G. Winter

Simulation as an Evluation Technique 10
Christine H. McGuire

- Essence of Simulation 10
- Techniques for Assessing Critical Components of Competence . . . 11
- Advantages and Limitations of Simulation 17
- Conclusion 23

Student Assessment at Alverno College 24
Georgine Loacker

Experience Based Learning and Assessment in Public School Systems. . . 33
Thomas R. Owens

- Introduction 33
- Assessment Problems 34
- Proposed Framework for Assessing Experiential Learning Programs . 38
- Recent and Emerging Developments. 49
- Bibliography 53

PREFACE

Educational measurement is currently experiencing a period of change and reconceptualization perhaps unprecedented since the days of Alfred Binet and James Cattell. As opposed to relying on traditional standardized objective academic achievement tests, many educators are searching for ways to assess learning that closely parallel real life tasks.

It is difficult to single out any specific contributors to the search for alternative measurement approaches, for they are all intertwined in subtle ways. A book, a speech, a legislative resolution, a new federal program, a newly developed test all play their parts in fostering change, reinforcing one another and generating new developments.

The purpose of the conference Alternative Conceptions of Competence Assessment was to present several new developments in educational measurement, contributing to a broadening of assessment options. A common thread running through the alternatives presented is their view of the complexity of human competence--and the inadequacy of the ubiquitous multiple-choice examination as a measure of competence. This underlying theme concerns the meaning of the term competence. As David Riesman (1979) has pointed out, society has until recently tended to regard competence in a narrow cognitive dimension symbolized by the phrase "scholastic aptitude," imposing technical mastery of subject matter as the standard of performance.

Reaction against the preoccupation with strictly academic verbal and quantitative skills is now evident in new kinds of competence valued by many educators: the ability to succeed in role-based, occupational or personal pursuits; generic abilities and interpersonal communication skills such as problem solving are highly related to the success of nonacademic tasks; and, in the affective domain, the development of self-confidence. Reconceptualizing the outcomes of education in these ways exerts a powerful influence on educational measurement. It would appear that curricular developments are showing signs of guiding the nature of testing, rather than vice versa. How quickly and how adequately the field of measurement can respond to new definitions of competence remains an open question. Much needs to be learned about tests of complex, generic, role-related performance before they are widely embraced as the proverbial "good thing." Herein we present an admittedly brief view of that response.

Conference Overview

The conference was divided into two parts, the first focusing on the conceptual basis for alternative competence measures, the second dealing with applications of nontraditional assessment techniques in particular contexts, for particular purposes. In actuality, this division blurs

because the speakers have illustrated their theoretical arguments with specific examples and have justified their applications on the basis of well thought out measurement philosophy.

The first speaker, David Winter, proposes a set of four characteristics that any competency measure ought to have, looking at a variety of assessment techniques in light of those standards. Dr. Winter is Director of Research at McBer and Professor of Psychology at Wesleyan University.

Next, Christine McGuire, Associate Director of the Center for Educational Development at the University of Illinois Medical Center, examines the use of simulation techniques in assessing four critical components of competence: observational and interpretive skills, problem-solving skills, interpersonal and communication skills, and technical skills. She devotes a major portion of her presentation to an analysis of the advantages and limitations of simulation as a measurement tool.

The third speaker, Georgine Loacker, describes the way Alverno College has redefined the nature and function of student assessment in a liberal arts curriculum. In a change process that has now spanned a decade, Alverno College has radically altered its view of the competent graduate and the way in which competence is demonstrated. Dr. Loacker is Professor of English and Chairman of the Communication Competence Division at Alverno.

Finally, Thomas R. Owens of the Northwest Regional Educational Laboratory reviews the difficulties inherent in measuring the outcomes of experiential learning programs at the secondary level. Here, outcomes are highly individualized, extending beyond the domain of strictly academic skills. He identifies a number of assessment strategies which take these difficulties into account.

What follows is likely to raise as many questions as are answered. We planned it that way, for by encouraging a questioning attitude among our audience, we hope they will start working on a few answers.

New Approaches to Competence Assessment

David G. Winter
McBer and Company
Boston, Massachusetts
and
Wesleyan University
Middletown, Connecticut

In my remarks today, I'd like to present several standards that any measure of any competence--especially an educational competence--ought to have. I will then discuss two or three traditional kinds of competence measures in light of these standards. Finally, I will introduce some new kinds of measures of competences and educational outcomes that we are working on at McBer.

(1) Any measure of an educational competence or outcome ought to look like it measures the things that it purports to measure. In psychological terms, it ought to have face validity. It may surprise you that I start there, because many psychologists would say that face validity is irrelevant. But if we seriously intend to use alternatives to traditional measures, either to certify students or to evaluate programs, these new measures must look like they measure the sort of thing they claim to measure. Faculty won't stand for anything else.

Let me give you a homely example. I believe my courses at Wesleyan increase students' interest in theoretical matters and perhaps decrease (if Feldman and Newcomb's review is correct) students' interest in religious matters. Now, a typical psychologist might try to measure this with an "item" such as the following from the Allport-Vernon-Lindzey Study of Values: "Would you be more likely to look at a newspaper headline that reads: (A) Scientist Announces New Theory, or (B) Theologian Discusses New Doctrine?" The thought behind the item is that

A is theoretical interest and B is religious interest. But if you seriously think that I, as a professor--especially if I don't have tenure--would stand for being evaluated by the percent of students who increase from choice B to choice A, you're wrong. Nor do I believe that questions of that sort will be appropriate for certifying my students. If competence measures are really going to be used in serious and important ways, therefore, they must measure the outcomes that they claim to measure at the same levels of complexity, subtlety, and sophistication at which those outcomes are taught. As a faculty member, I wouldn't permit any less. And that means a lot of good and popular traditional psychological tests won't be very useful for measuring competence in this way.

(2) A good measure should be generic, broad-based, and transferable to a variety of different subjects, disciplines and skill levels. We must avoid proliferation of measures. Otherwise we shall have an endless array of very specific, limited competencies, so that we worry about the equivalence of different measures of the "same" abstract competence.

(3) A good measure of an educational competence should be based on criteria which are public--criteria to which you can point and which other observers can use to arrive at the same evaluation. I don't just mean inter-scorer reliability in the psychometric sense. I also mean criteria which are public, so that I, as an evaluator, can justify my evaluation to a student who may think I'm capricious, or who may not understand why the evaluation was what it was.

(4) Finally, a good measure of any educational competence should have some empirically established validity as a predictor of important later-life, real-world outcomes. This is, of course, the familiar notion

of predictive validity or longitudinal validity, but it emphasizes predicting important later-life actions.

I don't think this series of four suggested standards is very startling or surprising, but I'd like to apply them to some old and some new ways of measuring competence. Let's begin with me as the Wesleyan psychology professor, using a very traditional means of evaluating my students' competence in something like critical thinking or analytic concept formation. I've given the following final exam essay question: "Compare and contrast the motivational theories of Freud and Maslow, with special reference to the issue of unconscious motivation." The act of evaluation, which is my grading the blue books, is probably going on late at night. I'm probably late getting the grades in and am under tremendous time pressure from the registrar. This is the fiftieth exam I've read, and everything is swimming in front of me. I pick up another blue book and I say to myself, "Hmm, this student says most of the right things about Freud and Maslow. The facts are pretty good, and the organization is all right, but it's awfully diffuse. The whole first few pages ramble on about Freud being linear and irrelevant to twentieth century America. It just isn't up to A standards. It's not sort of crisp--I can't quite follow it. But it's better than the usual B. Maybe I'll give it a B+. Well, is it worth an A-? Maybe I should go back and read it again. No, it's too late, and I have a lot of other exams to do. B+ it is." Maybe that's a slightly exaggerated version of what happens in my grading, and maybe other professors grade with a little more care, but it is a useful hypothetical example.

What about my question or measure, "Compare and contrast Freud and Maslow"? I think it has obvious face validity--it "seems" to measure the

things I think it should. I expect that most of my colleagues would agree with me that it's a reasonable measure, so it "passes" this first standard.

Is it public? Usually I don't even ask. I've never had another faculty member re-grade my blue books and see whether they agree with me, although I talk with other faculty members in a general way about whether particular students do as well for me as they do for them. I remember a very exciting moment once when one of my colleagues and I agreed that a particular student was the classic case of "a borderline B+/A-." We always had trouble deciding which grade to give, but in the end each of us had always decided to give him a B+. I guess you could say there is some agreement there, but the criteria of "A-" versus "B+" aren't really very public. I'm sure my students don't know what I'm grading <u>besides</u> just knowledge of facts about Freud and Maslow.

Is it generic? The question is confined to Freud and Maslow, but I believe that in principle it is generic because the same type of question could be used across a wide variety of subject areas. One need just fill in the blanks for whatever theorists were taught.

Does it have educational validity? I don't know. I would like to think that students to whom I give A's turn out as successes, and that my grading predicts professional success and personal satisfaction, but I don't check it very carefully and systematically. I don't really know whether my grades predict important things in later life.

Thus the traditional exam grading method of evaluating competence has face validity, has criteria that are public to an unknown extent (and therefore is of unknown reliability), has unknown relevance to later life, and is possibly generic. I tend to accept on the basis of faith or

a priori considerations that my measure of educational competence--the Freud/Maslow question and my grading of it--has all of these characteristics, but I am not sure and I never really check carefully.

What about an alternative, nontraditional measure in a competence based program? Suppose I don't ask students just to write exam questions, but rather ask them to engage in simulations, discussions, an in-basket exercise, or supervised medical or nursing rounds in a hospital. Suppose I ask them to produce talks or reports and then to assess their own product and themselves. These innovative and nontraditional means of assessment are attempts to break away from the traditional student testing.

I don't think that this change from old to new, in and of itself, solves any of the problems I've just mentioned. Face validity is still an issue. Faculty members in a competence based program will have the same agreements and disagreements and will need to discuss whether a given measure--whether it's a sample of behavior, a simulation, a product, a performance or a self-assessment--is an appropriate measure of the thing it purports to measure.

The criteria for evaluating the nontraditional measure (the product or the performance in the simulation) may or may not be public. At some places, like Alverno College, enormous effort goes into putting on paper the criteria used to grade or evaluate performance. I've had the great pleasure of visiting Alverno and seeing firsthand the detail, precision and care with which the criteria for competence assessment are specified, so that evaluators agree and converge in their judgments. That extra effort is necessary, because just calling a measure nontraditional or competence based or a behavior sample does not mean that it necessarily meets the public criteria standard.

Is the nontraditional assessment measure generic? Some are broad-based and applicable to many different fields, and some are not. Nontraditional measures are neither intrinsically generic nor intrinsically nongeneric.

Are they relevant to later life? A nontraditional measure is not automatically relevant to later life; it must be demonstrated to be so. I may have created what I consider a very realistic simulation of an important adult role performance, yet it is a simulation, and its predictability or validity with respect to performance in later life has to be established. If you think it doesn't need to be established, you are accepting it on faith, in the same way that I believe, without evidence, that my Freud and Maslow question is relevant to later life. Changing the labels, changing the names, and adopting new "fashions" therefore still leaves us with the same issues: reliability and validity (or as I would call it, public criteria and relevance), generic nature of the measures used, and face validity.

After that brief look at some old and new measures, I would now like to introduce some new measures. Several years ago, David McClelland, Abigail Stewart, and I became involved in a project to try to assess the effects of liberal education.[1] Now one of the things that seems to emerge in any statement of the goals of liberal education is a quality called "critical thinking" or "complex concept formation." Academics being academics, no one will agree with anybody else's exact definitions of these terms, but I think the general sense of the quality is clear. We would like to think that a liberally educated student can somehow go into a confusing, chaotic situation and create concepts that are useful

---

[1]Results of this project, supported by the Fund for the Improvement of Post-Secondary Education, are to be published by Jossey-Bass as Competence in College.

in organizing that situation, reducing the confusion to some kind of conceptual clarity. How are we going to measure this? The traditional psychological approach might involve some of the work on concept attainment, where the concept is defined in laboratory terms: blue circles over red triangles, some combination of shapes, or even (going back to Clark Hull) Chinese characters. Students are given examples that are and are not of the concept in question, and then asked to guess about new instances. The presumption is that the correctness of their guesses improves over time until they attain the concept. This measure involves concept attainment, however, and not concept formation. It doesn't seem anything like what I teach in any liberal arts class I ever teach. I don't use blue circles and red triangles, and I don't trust an evaluation at that level. In our project, therefore, we decided to take a new approach. If you want to measure students' abilities to make sense out of confusion, then throw them into confusion and ask them to make sense out of it. Typically, we have taken two groups of brief, imaginative stories (the kind of thing people write for a Thematic Apperception Test) and said, "What are the differences between these two groups of stories? Decide on the differences and write them out in any way, form, or style you like. You have half an hour." We call this the Test of Thematic Analysis.

By comparing the responses of freshmen and seniors at a liberal arts college we found that freshmen compare and contrast in chaotic ways. For example, they tend to make nonparallel comparisons and use their own emotional or subjective reactions as the basis of comparison. Seniors tend to use ordered, parallel kinds of categories. They talk about overarching issues which subsume the differences between the two groups of stories. Seniors manage to redefine and rework the categories to include

cases that may have initially seemed discrepant, and also to provide examples, exceptions, and qualifications.

Because we actually derived this measure from the differences between liberal arts freshmen and seniors, we know it possesses some educational validity. It reflects the sort of thing that seems to be learned in liberal arts colleges. You can substitute two ideas, two paintings, two novels, or two philosophical theories for the two groups of stories, so that it also seems generic. Scores for this measure are based on a series of category definitions that can be taught well enough that inter-scorer reliability coefficients of .85 to .90 are achieved. There's no reason why students couldn't be taught how this measure is scored. Indeed, it might be the basis for teaching them how to write better.

We are now exploring the real-life relevance of this skill. That is, if you score high in Thematic Analysis, how is it relevant to adult world-of-work role performance? One preliminary finding, interestingly enough, is that good alcoholism counselors in the Navy ("good" in the sense that their clients get better) are better at this sophisticated concept-formation ability than are not-so-good alcoholism counselors. That is just one finding, but it illustrates how one might go about establishing that being able to draw these sophisticated, precise, articulate kinds of comparisons and contrasts is related to real-world skill and performance.

I will mention briefly another measure being developed at McBer. In some other research for the Navy, we kept finding that some kind of general skill at planning, organizing and coordinating different tasks--so that they're all done at the right time and in the right sequence--seemed to be important for successful leadership. I think that this is a skill that is learned in college, too. You have to get term

papers, lab exercises, exams and everything else done; you have to maneuver your time to fit things in without any waste. It's a skill that I also think is important in getting a gourmet dinner on the table at the right time--with the right sauce over the right meat, neither too cold nor too hot, with enough batches of rolls freshly done, the table is set, and everything ready at the end.

We have tried to simulate this--to create a brief and abstract test of this skill. Several different subtasks, lettered A, B, C, D, and E, must all be accomplished with the greatest possible efficiency. Associated with each task is a time and a set of rules; for example, task A has to be done before task C, task D has to be done three times with intervals of five minutes between each, tasks B and C have to be done together, and so on. The point of the test is to get the whole series of tasks done as soon as possible, with as little waste time as possible.

We found that high scores on this test "Scheduling Exercise" predict superior performance among Naval officers, especially at the very top levels in the chain of command. We are now trying it out in an educational context, seeing whether this skill increases as a result of certain kinds of liberal arts training. It's clearly a publicly scorable measure (a computer can score it, actually); it seems generic, and I hope it has face validity.

McBer is a company that is committed to developing a variety of measures that satisfy the four criteria I mentioned. The two examples I have described illustrate some specific measures of this sort. While they may also be relevant to measuring competence in higher education, my real concern is to emphasize the importance of the four standards, for any measure--traditional, "competence based," traditional psychological test, or innovative new assessment procedure.

Simulation as an Evaluation Technique*

Christine McGuire
Center for Educational Development
University of Illinois at the Medical Center

Some more or less exotic simulations and games have long enjoyed a deservedly favored status in the training of business executives and military personnel. However, it was World War II that gave the main impetus to systematic development in the theory of gaming and simulation as techniques for use in training and instruction. In the newer and more hazardous occupations--that grew so rapidly in that period--it was recognized from the beginning that training and testing in the real world were either too risky--as in the case of commercial pilots, or were simply unfeasible--as in the case of astronauts. It was this recognition which gave rise to extensive simulation for instruction in those fields. However, the principles of simulation which underlie the now-familiar management games, military exercises, space and aircraft simulators have only recently been introduced in medical education. In this field the potentialities of simulation not only in the instruction, but also in the assessment, of physicians and other health professionals are only now beginning to be appreciated.

Essence of Simulation

Perhaps the limited exploitation of these potentialities is in part due to the vision of astronomical costs conjured up by the word

*This paper has also been presented at the 1976 Annual Invitational Conference of the National Board of Medical Examiners which was funded in part by the Henry J. Kaiser Family Foundation and the Commonwealth Fund. Proceedings of that conference have been published in the volume, An International View of Qualification for the Practice of Medicine.

"simulator." Yet, reduced to its essence, a simulation designed for instruction or assessment consists simply in putting an individual into a realistic situation and documenting the individual's performance. It is widely known that the best prediction of future performance is present performance in a closely similar situation, and it is this principle which suggests the potentialities of simulation in the assessment of professional competence. Exercises designed to exploit these potentialities are characterized by four essential features: (1) the test situation imitates (does not duplicate) some aspect of reality; (2) it requires the active participation of the respondent; (3) the respondent's participation triggers appropriate feedback which may or may not modify the situation but which can, in any case, be utilized for subsequent decisions about pending action, and (4) this action of the respondent may modify the problem.

In continuing experimental work at the University of Illinois, the University of Southern California, Michigan State University, McMaster University, National Board and many other places, this essence has been captured in a variety of simulation modalities that are economically and technologically feasible to use in the assessment of professionals.

Techniques for Assessing Critical Components of Competence

Data are beginning to accumulate that simulation is particularly useful in the assessment of four critical components of competence expected of every health professional: observational and interpretive skills, problem-solving skills, interpersonal and communication skills, and technical skills.

Assessment of Observational and Interpretive Skills

In the assessment of observational and interpretive skills it is clear that provision must be made to test the physician's ability to interpret data that may be received through different sensory modalities: visual data, auditory data or data from palpation and percussion, to name but a few. Consequently, to assess these important skills, simulated clinical and laboratory data are presented to examinees in as nearly realistic form as possible and the examinee is simply asked to note and/or interpret them.

Exercises of this type may be based on photographic reproduction of x-rays, gross and microscopic specimens, lesions, tracings and the like--all presented in a form that imitates life as closely as possible. Sound simulators developed for playback of high-fidelity recordings of heart, breath and abdominal sounds through individual stereophones may be employed for auditory data. Patient findings that involve a combination of sound, color and movement may be presented in high quality movies or videotapes depicting the physician-patient interview and/or certain aspects of the physical examination. More recently, several three-dimensional models have been developed; varied pathologies can be inserted in these models that require the physician to demonstrate both skill in eliciting the findings (see below) and ability to interpret them.

Using any or all of these modalities, the individuals being tested can be required to demonstrate that they can describe precisely the findings presented in auditory, visual or other forms, can interpret the significance of these findings, can anticipate other findings that might be related to those demonstrated, can predict what might exacerbate or

relieve the symptoms, and can make some initial judgments about their significance and interrelationships.

Assessment of Problem-Solving Skills

But observational and interpretive skills comprise only a few of the many elements requisite to effective problem solving, and only limited aspects of the total competence that is denoted by the term "clinical judgment." Thus, in the testing of physicians, it is necessary to try to measure as directly as possible the individual's skill in dealing with problems in patient management. Indeed that may be the heart of the task, and the part of it for which simulation is especially valuable.

Any patient management problem that purports to simulate the physician-patient encounter must have the following characteristics: first, it must be initiated by information of the type a patient gives a physician, not by a predigested summary of the salient features of the case, and if it is to be realistic, it must be couched in terms which the patient or a referring physician would use. Second, the exercise must require a series of sequential, interdependent decisions representing the various stages in the diagnostic workup and management of the patient. Third, the examinee must be able to obtain information in realistic form about the results of each decision, as a basis for subsequent action. Fourth, once these data are obtained, it must be impossible for the examinee to retract a decision that is revealed to be ineffectual or harmful. Fifth, the problem must be constructed so as to allow both for different medical approaches and for variation in patient responses appropriate to these several approaches. Hence, provision must be made for modifications in the problem as the patient responds to the specific

interventions chosen by each examinee. Finally, it is essential that these modifications, including the complications that develop and must be managed, differ from candidate to candidate according to the unique configuration of prior decisions each has made.

Problems meeting these criteria have now been developed in three modalities for use with medical students, residents and clinicians: (1) a paper and pencil format using latent image or opaque overlay techniques for feedback systems, in problems which are suitable for either individual or group administration and are amenable to computer scoring and analysis;[1] (2) various computer formats, some of which employ unconstrained natural language and all of which are designed for individualized administration to untrained subjects; (3) an oral interaction format suitable for role-playing exercises in individual or small group settings.

There is also, in principle, a fourth modality for presenting simulated patient management problems--the computer-managed robot--which may ultimately be the most effective of all. The most elegant of these--SIM I, developed at the University of Southern California for the training of anesthetists--can be programmed not only to display various combinations of findings, but also to present a number of problems and to respond appropriately to different physician interventions. Indeed, SIM I will do everything including vomiting on the physician who gags him

---

[1]For examples of such simulations see McGuire, C., Solomon, L. and Forman, P. (Eds.) Clinical simulations (2nd ed.). New York: Appleton-Century-Crofts, 1976; for instruction in building various types of simulations see McGuire, C., Solomon, L. and Bashook, P. Construction and use of written simulations. New York: The Psychological Corporation, 1976.

when inserting an endotracheal tube. The robot may even "die" if the physician manages him inappropriately or responds inadequately to changes in his condition. However, the lovely thing about this patient is that he never gets tired of being examined or treated and he can be revived as often as required in the training and assessment of anesthesiologists.

In short, written simulations, live simulators, computer simulations, and automated robots can all be used to present clinical problems in a realistic form that requires examinees to respond as they would in the office or clinic setting, to receive life-like reactions from the simulator and to cope with the patient's responses to a sequence of inquiries and interventions. Challenging problems presented in this form have the emergent, dynamic quality and feedback capability characteristic of real-life settings.

Assessment of Interpersonal Skills

A third component of competence of increasing interest is the area of interpersonal and communication skills. Building on the work of Barrows and his colleagues,[2] some of the most promising developments for assessing these skills take the form of live simulations of interviews and conferences in which the physician-patient or physician-colleague interactions can be objectively documented. These simulations differ from live simulations of clinical problems only in their focus: in the former, the emphasis is on solving a complex medical problem; in the

---

[2]See Barrows, H. Simulated patients. Springfield, Illinois; Charles C. Thomas, 1972, and Levine, H. and McGuire, C. The use of role-playing to evaluate effective skills in medicine. Journal of Medical Education, 1970, 45, 700-705.

latter, the situations are chosen to put stress on candidates' communication skills, their sensitivity to patient and colleague needs and their attitudes toward these important people.

For example, in what purports to be a straightforward diagnostic interview, the simulated patient may, if appropriately handled by the physician, provide cues that the real problem is a deep-seated personal, familial or emotional one. Examinees' sensitivity and response to these cues and their skill in handling such a patient, and in gaining the "patient's" understanding and cooperation in a proposed plan of management, are readily assayed in such simulated interviews.

Analogous simulated interviews with colleagues have been developed to deal with referral and consultation requests and even with those ticklish situations in which colleagues differ sharply about the management of a particular patient. Finally, the simulated interview may be designed to require candidates to communicate with one or more members of the health team. For example, they may be required to give instructions to a simulated nurse or to request assistance for a patient from a simulated dietician or social worker, or to make a presentation and respond to the reactions of invidivuals who have been programmed to take the role of other members of the health team in a simulated staff conference. In all of these situations it is possible to obtain objective ratings of candidates' communication skills, interpersonal skills, attitudes and even their professional demeanor in lifelike settings.

Assessment of Technical Skills

In order to test the physician's technical skills in a simulated setting, it is necessary to have access either to a very large number of

paid patients whose condition is stable and who present varied pathologies, or to three-dimensional models that can be made to simulate different pathologies in some kind of lifelike object which the physician can examine. A few such models now exist--for example, the larynx, ear, intubation and eye models--into which varied pathologies can be inserted. With these models students can be required to demonstrate that they have mastered skills of using sophisticated instruments both to collect data and to interpret them. While offering great promise as a means of objectively testing technical skills at all levels, the presently available models are still too crude for use with advanced students or practicing physicians.

Advantages and Limitations of Simulation as an Assessment Technique

Table I summarizes the types of simulation now available for use in testing health professionals. In considering the relative merits of these techniques, one must ask about their advantages and limitations compared to what and for what purpose. In such discussions, many thoughtful people are inclined to view the main advantage of simulation as deriving from its relation to reality; such persons are then likely to conclude that reality (i.e. actual patients) would serve still better as "testing material." In my view this is not the case. For most purposes I regard simulated reality as a more effective device for sampling physician behavior than either conventional testing methodologies or reality itself.

Simulation imitates, it does not duplicate, life; this is both its greatest strength and, in the minds of some, its greatest weakness.

Clearly, there are some aspects of reality that cannot at the present time be economically simulated, if they can be simulated at all. Thus, it is necessary to be very cautious about predicting on the basis of responses to a simulated situation how an individual will behave in reality; at best it is possible to predict how well that individual is capable of behaving.[3] Secondly, it is important to recognize that simulation is not an appropriate method for testing all aspects of performance. For example, simple recall of factual information is more economically and directly measured by conventional techniques of objective testing. At the other extreme, professional habits can be assayed only by careful and repeated observation over a long period of time in diverse settings. Between these two extremes simulation provides the following important advantages.

Perceived Relevance

As compared with more conventional evaluation strategies, simulation offers the possibility of designing problems so that they correspond more closely to the life situations a student actually faces. Tests and summative evaluations composed of such problems appear to both candidate and student as far more relevant than typical textbooks or conventional multiple-choice examinations. Such perceived relevance is at the very least psychologically beneficial. And, in these days of concern about the cultural biases of our tests, perceived relevance may turn out to be our only legal defense.

---

[3] The same caution must, of course, be observed in predicting how individuals will behave in a new life situation on the basis of how they behaved in a prior, necessarily somewhat different, life situation.

TABLE 1: SIMULATION AS AN ASSESSMENT TECHNIQUE

Competence Tested	Type of Simulation	Illustrative Techniques
Observational and interpretive skills	Simulated clinical and laboratory data	Photographic reproductions Sound simulators Movies and video-tapes Three-dimensional models
Problem-solving skills	Simulated problems in patient management	Written simulations Live simulations Computer assisted simulation Automated robot simulations
Interpersonal and communication skills	Simulated interviews and conferences	Live simulations ("programmed" patients, colleagues or teams)
Technical skills	Simulated pathology	Three-dimensional models

Predetermination and Preselection of the Task

It is also of prime importance that by utilizing simulation techniques this perceived relevance can be achieved without being dependent on the accidents of nature and the flow of real problems available at the particular place and the specific moment in time when an assessment is to be made. It is obvious that simulation makes it possible to predetermine precisely the exact task which examinees are to be required to perform. Further, it is clear that, in contrast with the "noise" always present in reality, simulation makes it possible to focus on the elements of primary concern in a testing situation and to eliminate irrelevant and confusing complexities that would contaminate the assessment. In addition, the fact that assessment tasks can be preselected and predesigned means that they can be carefully graduated in difficulty to be appropriate to the precise group with whom we are concerned. By developing a standard set of parallel simulations it is possible to confront candidates over and over again with interesting variations of what is essentially the same task, until there is clear evidence that each has mastered it.

Standardization of the Task

Just as a given student can be repeatedly confronted with the same task, simulation enables an examining body to standardize the task for all examinees and, in the case of medical examinations, to do so without subjecting one or a few patients to repeated harassment by a large number of candidates. In short, all examinees can be given exactly the same problem to cope with and this can be accomplished without an attack on nature.

Improved Sampling of Performance

By standardizing the tasks and focusing on the most significant aspects in each it is possible in a given time period to sample an individual's performance with respect to a much broader and more representative group of problems. This, of course, requires careful selection and definition of the tasks, but that is precisely the luxury simulation affords, which reality can rarely provide in a reasonable time frame.

Furthermore, if you wish to test the candidate's ability to respond to an emergency or to handle chronic disease, it isn't necessary for either the examinee or you to wait years for the "right" emergencies, or to watch the gradual deterioration of real patient over a lifetime. In carefully developed simulations any problem, ranging from the most urgent emergency to illness spanning many years, can be collapsed into a half-hour exercise and summoned on demand.

Increased Accuracy of Performance Ratings

When the exact tasks that are to compose an examination are precisely defined and preselected it is possible to develop specific, detailed criteria for judging examinee performance, and to train examiners in applying these criteria equitably, consistently and in a manner to provide feedback of maximum benefit both to the candidate and to those who must judge his or her eligibility for licensure or certification. Because this type of training of examiners is feasible, there is no question that a higher degree of interrater reliability can be achieved in the scoring of examinee performance on simulated problems and simulated interviews than on real problems with real patients.

Beneficial Byproducts

Increased Responsibility and Realistic Feedback in a Practical Time Frame. One of the most important advantages of simulation over reality in my view consists in the fact that all candidates can be allowed full responsibility to careen down their merry way to disaster without any risk whatsoever to anyone or anything other than a piece of paper, a computer or the psyche of an examinee. This is important not only in teaching but also in testing. Subject to the non-so-tender mercies of a neophyte, a simulated patient can become progressively sicker, can develop iatrogenic complications, can commit suicide, can die "naturally," and still can be resurrected tirelessly to serve again in the retesting of the same or other candidates. And at every stage of the evolving problem any examinee can be provided with instructive feedback about the patient's response to each intervention.

Increased Learning. Further, the nature of the feedback provided in simulations as compared with that available in either conventional testing or reality should be noted. While the former rarely provides any genuinely relevant feedback, the latter often furnishes it only indirectly, equivocally and after long delay. In contrast, the prompt, specific and unequivocal feedback characteristic of well designed simulations makes them a powerful tool not only in assessment, but also in the enhancement of learning. Surely this is an important consideration in the choice of any evaluation strategy.

Educational Impact. Finally, one of the main advantages of simulation is its educational impact. Tests are powerful instruments in directing the attention and efforts of thousands of students, house

officers and, in the future, perhaps physicians in practice, toward whatever is regarded as important in their licensure and certification. Effective simulations have the power to direct that attention toward developing those aspects of competence which the medical profession claims to treasure, and, as we have seen, the feedback that occurs in the simulated situation is, itself, a tremendously powerful tool for improving performance.

Conclusion

These, then, are the oppportunities and advantages which well constructed simulations offer. They provide these benefits in what is one of our most important tasks in health profession education--assessing the performance characteristics of those individuals who have responsibility for the health care of our population.

Student Assessment at Alverno College

Georgine Loacker
Alverno College

In presenting assessment at Alverno College, what I'd like to do is explain how we developed it, how we define it, and how it works. In doing so, I'm describing a process we developed without most of us even knowing at the time that there was a movement in competence based education. The process began with our looking at our students and saying "What can we do for them that we aren't doing now--and would like to do further?" So, I'm going to begin with how we developed our curriculum because the development of assessment follows directly from that.

I'm sure you're not terribly interested in the historical development of the curriculum, but I will talk about it a little bit to make sense out of what follows. First of all, Alverno College is a private women's college in the Midwest. The enrollment is between 1200 and 1300 students, a little over one-third of whom are older, nontraditional students from 26 to 65. Much of the student body is made up of "first generation" college students. Seventy-five percent of the students receive financial assistance in pursuing their education. I mention that because often when we talk about a small, private college, people get the idea that it's made up of students who can afford a very elite school; but we are particularly concerned with the education of the middle class woman. We've always had a tradition at our college of education rather than research. We're interested in what we can accomplish with our students--not in doing research on them. However, since we began this program, we've been engaged in an increasing amount of research to improve the educational process.

In the 1960s, we realized we had to expand the lay student body if we were going to justify our existence as a college. Even though there were fewer and fewer religious sisters to be educated by us, we decided that we had started such a good thing that what we were doing was worth expanding. However, we said to ourselves, "We don't deserve to be the ones to do that unless we can accomplish something that cannot be accomplished elsewhere. There are several good colleges and universities in the city and there's no point to our existence unless we can provide something for our students that is a real alternative to what they might receive elsewhere." So in 1969, approximately, in our continued probing of the educational process to find out how we could improve it, we began to explore what that alternative might be. By 1971, we were focusing on some questions that our administration had addressed to us. We had a new president at that time, and (because new presidents have at least one year of grace) she suggested that the liberal arts departments meet with each other and justify their courses of study to each other. "Why should anyone study with you in the art department?" "Why should anyone be an English major?" "What course are you teaching that no one can afford to miss--and why?"

In the process of questioning one another's disciplines, the faculty made a decision to at least look into the possibility of defining a degree in terms of outcomes. By 1972 we had identified four outcomes: involvement, problem solving, communication, and valuing. For a whole year we probed those outcomes to see if we could agree that they might be the outcomes of a liberal education. We argued a lot about that issue; possibly because we couldn't picture defining all the wealth of education in four outcomes. We eventually made a decision to go with outcomes when we were able to expand them into eight.

During the year of 1972 to 1973, the administration established an academic task force to take the outcomes that the faculty had agreed upon and turn them into a curriculum. They, in effect, became the staff to the faculty, and created the curriculum that is now in existence. The first five outcomes or student abilities were (1) communication skills (including reading, writing, speaking, listening, quantitative literacy and media literacy), (2) analytical capabilities, (3) problem-solving ability, (4) valuing in decision making, and (5) social interaction skills. The last three are clearly more related to specific content: (6) understanding of the environment, (7) understanding of the contemporary world, and (8) aesthetic competence (understanding and responding to the arts and humanities).

We had great long and interesting arguments about "How in the world can you evaluate valuing? How in the world can you evaluate aesthetic competence?" One response was that art teachers are always evaluating aesthetic competence at the end-of-year exhibits of seniors. If we could just say to each other what the basis is for making those judgments, we could begin to develop some criteria that would be worth looking at. Such criteria would tell our students what we are looking for.

In designing the curriculum, we decided on a pedagogically developmental model, breaking each outcome or competence into six levels. Why six? It was not a sacred number, it was an arbitrary number that we could work with. We applied no particular taxonomy or developmental philosophy; we broke open each of the competences into levels in a slightly different manner. As a faculty, we took our courses and decided which of the competences seemed natural to our disciplines. We then translated each course into an integrated expression of these competences so that students would have a place to learn them.

We were operating on the basis of the following assumptions: First, we decided that education has to go beyond knowing, that students have to be able to do something with their knowledge. Second, we decided that our competences would be stated as generic abilities. We didn't want an endless list to suggest what one needs for graduation; we have lists, but those lists further specify what we're trying to do within our own disciplines. Third, we agreed that competence can be developed. It's the luxury of a lot of researchers to say, "Well, maybe these competences cannot be developed." But since our business was to develop them, we had to assume that they could be developed. When people would say "Are you sure?" we'd simply say, "No, but we're trying to find out. And since no one else has the answers, it's a lot more fun to try to find out. At least we'll end up with students that are more competent than they were before." Fourth, we assumed that competence can be assessed. Again, we realized it would take a lot of work before we could even say that maybe we're beginning to touch what that competence might be.

Several implications followed from this last assumption. The first was that competences must be carefully identified and compared to life performance. Thus, our Office of Evaluation began to investigate the validity of our assessment instruments by looking into the match between competence required in real life roles and those we defined. A second implication was that if we're not going to test only knowledge but what people do with knowledge, we would need to sample actual behavior. Assessment would have to involve the sampling of behavior that is then very carefully observed and judged on the basis of criteria. Those criteria would have to be made public. We would tell a student beforehand what she is expected to demonstrate. We also realized that learning would have to be active--that in addition to lectures and other

kinds of input, students would need performance models from their peers and the faculty. A final implication was that the competences must be developed and assessed in multiple modes and contexts. Because one sample is so different from another sample, we don't know how many different kinds of samples one needs before one can infer a general competence.

As we define it now, assessment is essentially sampling, observing, and judging on the basis of explicit, public criteria. We recognize the fact that a person is infinitely complex and infinitely varied and that the behaviors by which that person expresses his or her personality are just as complex and varied. When we take a sample, it is just a little piece on a given day. As a result, we try to take as many samples as we can, in different contexts, with different kinds of instruments, with different assessors. What that sounds like is that our educational institution has become a multiple assessment center, which, in effect, it has. Beside the need to sample behavior on different days in different contexts, we also realize that we could apply different kinds of criteria to a single sample. For instance, if we have a sample of student performance, we might not only be able to apply critical thinking criteria to it, but we might also apply communication criteria and judge whether the student was a good communicator. We might also judge that same sample for the problem solving behavior displayed. All the assessments that we do can be interconnected, reused, and used for different purposes, assuring transference and reinforcement of the behavior as the student moves from discipline to discipline, class to class, and so forth.

As a result, we've discovered that our assessments are generic, developmental, and holistic. Like the competences themselves, our

assessments are generic, rather than task-specific. In order to be generic, they must be both external and multiple. In order to be developmental, the criteria need to be explicit, and assessment results must be presented to the student immediately and in detail. We've learned that one of the best techniques for the development of competence is assessment itself, through the careful use of feedback. Finally, assessments are holistic--i.e., both integrated and open-ended. That is, we try to create assessments in which a student is not limited to demonstrate only what we're looking for, but in which she can demonstrate higher levels of a competence. And though at early stages we look at components separately in order to make competence manageable for learning, at advanced levels we look at several competences operating in concert.

Thus, the assessment process as it operates at Alverno--whether the student is assessing herself or whether the faculty or a peer or another professional is assessing a student--is viewed in terms of this constant process. We identify the outcome and define the criteria. Performance is then sampled, recorded, and carefully judged, and feedback is given which (if the goal is not achieved) leads back into additional learning experiences for the student.

The kinds of assessment techniques that we've created include what we might call proactive paper-and-pencil tests. We use that word to distinguish them from the multiple-choice or any kind of forced choice format in which students are given answers to select from. In such instruments, students create their own answers. Often when we talk about competence, people say "Well, you won't have any paper-and-pencil tests." We have hundreds and thousands of paper-and-pencil tests, but we aim to make them as proactive as we can so the student can really create her own

answers. We also use a lot of work samples; we try to catch the student at work, particularly if she's doing an internship or any other kind of experiential learning. We use in-baskets, a type of written simulation which some of you are familiar with. We use a lot of role playing, particularly in oral and group simulation situations. We use self-assessment as one of the most important means to learning, because we believe that students cannot become lifelong learners until they can evaluate themselves. If they're always dependent on someone outside themselves to judge whether they are doing a good job or not, how can they become independent learners? Thus, we've committed ourselves to assist them in developing their own ability to assess themselves.

Let me give you a few examples of how the assessment process actually works. Seventy-five percent of the assessment is done within our courses. The faculty design and administer the assessment instruments for those courses. However, some of the assessments are totally external to courses and--although designed by the faculty--are administered by our assessment center. When students enter Alverno, for example, they take the first stage of the communications assessment. They have to give a speech, they have to write a letter, they have to listen to a videotaped lecture and answer open-ended questions about it, and so forth. Trained paraprofessionals assess all of these samples on the basis of specific criteria.

The following excerpts from the instructions provided to students engaged in the communications assessment illustrate the nature of the exercise.

> Speaking (Form A) - Part 1. "Should college be open to anyone who wants to try it or should there be some criteria operative to screen students before they enter college?" Take a position on either side of this question and prepare a 3-5 minute speech to convince an audience of the worth of your position. Your

audience will be a group of high school students who are looking into their own futures. Use specific instances from experience and observation to support your ideas and clarify them for this audience.

Media (Form A) - Part 1. After you have your notes organized and are ready to do your speaking prepare a visual aid to incorporate into your presentation. It can be a diagram or mini outline or any other kind of illustration. It should assist you to make your point and assist the audience to understand your message. Use the newsprint and pen provided. Take the finished product along to the video studio.

Following these communication exercises, students are taught to assess themselves. They observe their own speaking performance on videotape and then are asked a series of questions and must rate themselves. After they assess themselves on each of the criteria, they meet with the paraprofessional who gives them feedback on their performance in all the areas that I mentioned, comparing it with their own ratings. Thus, students are "validated" on the basis of their ability to recognize their own strengths and weaknesses.

Another example of external assessment is found in assessing understanding of the contemporary world. Students come prepared to discuss an issue in a group setting. Because we feel that adults demonstrate their knowledge of the contemporary world in discussion situations, we're willing to "muddy up" the testing situation with a discussion technique, even though we're looking for understanding of contemporary events. The exercise is described as follows:

> Assessing a given student's achievement of Level 1 of Contemporary World will involve the close observation of that student in a structured group interview/discussion session with two other students and an interviewer. The student prepares for this interview session by choosing a particular perspective from which to discuss an energy-related issue. As the interviewer directs the students to identify and explain significant international, national, and local events related to the issue, the assessors will judge whether or not a given student fulfills the above criteria.

To assess performance in this kind of situation, we use a team of assessors typically made up of people from the Milwaukie business and professional community, advanced students, and faculty--all of whom are trained to evaluate students on the basis of specific criteria. Professionals and business people in the Milwaukee community volunteer their services. And this is one of the most lucky and successful things we've hit upon. We never dreamed that business people would be eager to get inside our ivory tower and contribute by assisting with the evaluation of students. We're finding that they are some of our most faithful and loyal assessors.

Earlier there was a reference to the great effort it takes to accomplish something like this with students. The question I would like to address is, "Is the effort worth it?" Right now there's no way for me to answer it except very subjectively. Our students are our best answer, and we're excited when we and other people interview them. Ultimately, our Office of Evaluation will demonstrate whether or not the effort is worth it from the students' point of view.

The last thing I would like to insert is something people don't often remember--how it's worth it for the faculty. What we've discovered is what the Alverno learning process has done for ourselves. We've always been concerned with the student, but this development has forced us to move out of ourselves into concern for what the student is doing--to look at what can I do to assist the student, rather than what do I know that I can give the student. This focus has been drawn out of us by the entire process. As a faculty member, even if that was all it accomplished for my future--to set me on that track for the rest of my life--I would say it has been worth it.

Experience Based Learning and Assessment in Public School Systems

Thomas R. Owens
Education and Work Program
Northwest Regional Educational Laboratory

Introduction

Since 1970, seven national panels have reviewed American secondary education. In each case, one of the major concerns expressed by these panels has been the tendency for our schools to keep young people from having educational experiences which lead to the development of a sense of adult responsibility. In response to these challenges and the desire of many young people to engage in meaningful learning environments, hundreds of experience based learning programs have emerged throughout the country in both public and private schools. Often these are referred to as experiential learning programs. As used in this presentation, experiential learning refers to a process of learning gained from both planned and unplanned experiences which involve youthful learners in meaningful activities and relationships with adults in the community. The learner is helped by another person to examine the meaning and implications of these experiences for his or her future growth. James Coleman (1976) contrasts action-based experiential learning with "knowledge assimilation" in which a teacher passes on a fixed amount of knowledge to students.

In a recent review of evaluation findings of experiential learning programs done for the Office of Assistant Secretary for Policy, Evaluation and Research, U.S. Department of Labor (Owens, Owen and Druian, 1979), programs were classified into three types: work

experience and training programs, career or vocational education programs and non-career-oriented experiential programs. These non-career-oriented programs take many forms, including community improvement, service learning, outdoor education and cultural journalism. Although some experiential learning programs focus on a single student population such as low income students or gifted and talented students, others such as Foxfire or Experience Based Career Education have been developed to appeal to the broadest spectrum of student backgrounds and ability.

Today I would like to share with you my perceptions of some assessment problems faced by experiential learning programs, a few possible solutions and a hopeful view of newly emerging developments that are designed to improve the evaluation of experiential learning.

Assessment Problems

To avoid presenting a laundry list of problems faced in assessing experiential learning, I propose that we consider these problems from a psychological, conceptual and operational perspective.

Psychological Problems

Psychologically, many program staff involved in experiential learning programs are reluctant and sometimes fearful of having their activities assessed. In many ways their fears are well founded. During the late sixties, at the height of the accountability movement in education, there was strong pressure for external evaluations of new programs to assure the taxpayers and funding agencies that their money was being wisely invested. In an attempt to achieve simple cost-effectiveness ratios, program outcomes were often assessed along only one or a few dimensions relying especially on standardized achievement tests or use of indicators

such as the drop-out rate. Such practices not only magnified the weaknesses inherent in reliance on standardized tests but they caused educators to become suspicious of evaluators. Educational practitioners began to view the evaluation process only as an external mandate rather than as a process that could be an effective tool for helping them better understand and improve their programs. The accountability movement also reduced the initiative of many staff in conducting their own informal evaluations since they saw the outcomes of evaluation more in terms of risk and harm than in terms of positive feedback or constructive use.

Conceptual Problems

Conceptually, assessment of experiential learning programs has been hindered by four areas of inadequacy: 1) clarity regarding students' outcomes; 2) appropriate strategies for helping to achieve these outcomes; 3) relationships between short- and long-term outcomes, and 4) a theoretical or research framework in which to view the outcomes.

The Experiential Education Guidelines developed last year at the National Center for Research in Vocational Education (Miguel, 1978) contains a set of 16 guidelines proposed by a national advisory panel involving business, labor and community members. Their first guideline addresses the need for clear learner objectives. These guidelines cite three primary reasons for lack of clarity in stating objectives: "(1) difficulty in stating some of the learner outcomes; (2) preference for flexibility; and (3) uncertainty in program design or execution." This advisory panel identified four potential benefits of clear learner outcomes: "experiences can be targeted to achieve objectives; experiential learners and persons interacting with them will have an understanding of their roles; learners will engage in activities

purposefully and experiences will be meaningful and productive; programs can be evaluated properly."

Many evaluations of experiential education programs attempt to assess short-term student outcomes without relating them to longer term benefits beyond the school years or to a theoretical or research base. As a result, even when short-term findings indicate significant student growth, there is no basis for interpreting these findings in terms of a larger framework. Failure to carefully assess program treatment strategies prevents evaluators from relating student outcomes to particular learning processes used in the program.

Another conceptual problem involved in assessing experiential learning programs is related to some essential characteristics associated with experiential learning. Such programs are usually individualized so that not only are students using different experiences to reach some common program objectives but often the objectives themselves are individualized. Thus, some students will be engaged in experiences leading to one outcome while other students may have selected other outcomes to work toward. Associated with this individualization of student objectives and learning strategies is the fact that another important ingredient of experiential learning acknowledges that students, while learning in the community for example, will identify and gain some important outcomes that were not anticipated at the time they prepared a project or learning contract. These situations call for a different approach to assessment from that which may occur in a classroom after a teacher has spent a week teaching the entire class prespecified factual information about a particular period of American history.

Operational Problems

A number of authors have given considerable attention in recent years to the problems of implementing an evaluation. Therefore, within the context of this paper, I wish to mention only three problems--those of selecting variables, research designs and appropriate assessment techniques. In experiential learning programs student learning objectives tend to focus less on knowledge of factual content and more on applying higher level thinking skills, on attitudes and on personal and social development skills. Thus, the delineation of what project staff mean by good work attitudes, a sense of personal responsibility or decision-making skills becomes important and is time-consuming. This leads to the need for an evaluator to be selective in identifying a modest number of outcomes to assess.

Recognition of the fact that different students within an experiential learning program may be working to gain unique learner objectives rules out the appropriateness of most experimental research designs which assume all students in a program are receiving common instruction that is not being taught to those in a control group. Thus the selection of evaluation designs that are sensitive to the unique characteristics of experiential learning programs is an important but often neglected aspect of evaluation. The unique learning objectives of individual students coupled with the vagueness of some of the statements of learner outcomes make the job of selecting or developing appropriate assessment techniques particularly difficult.

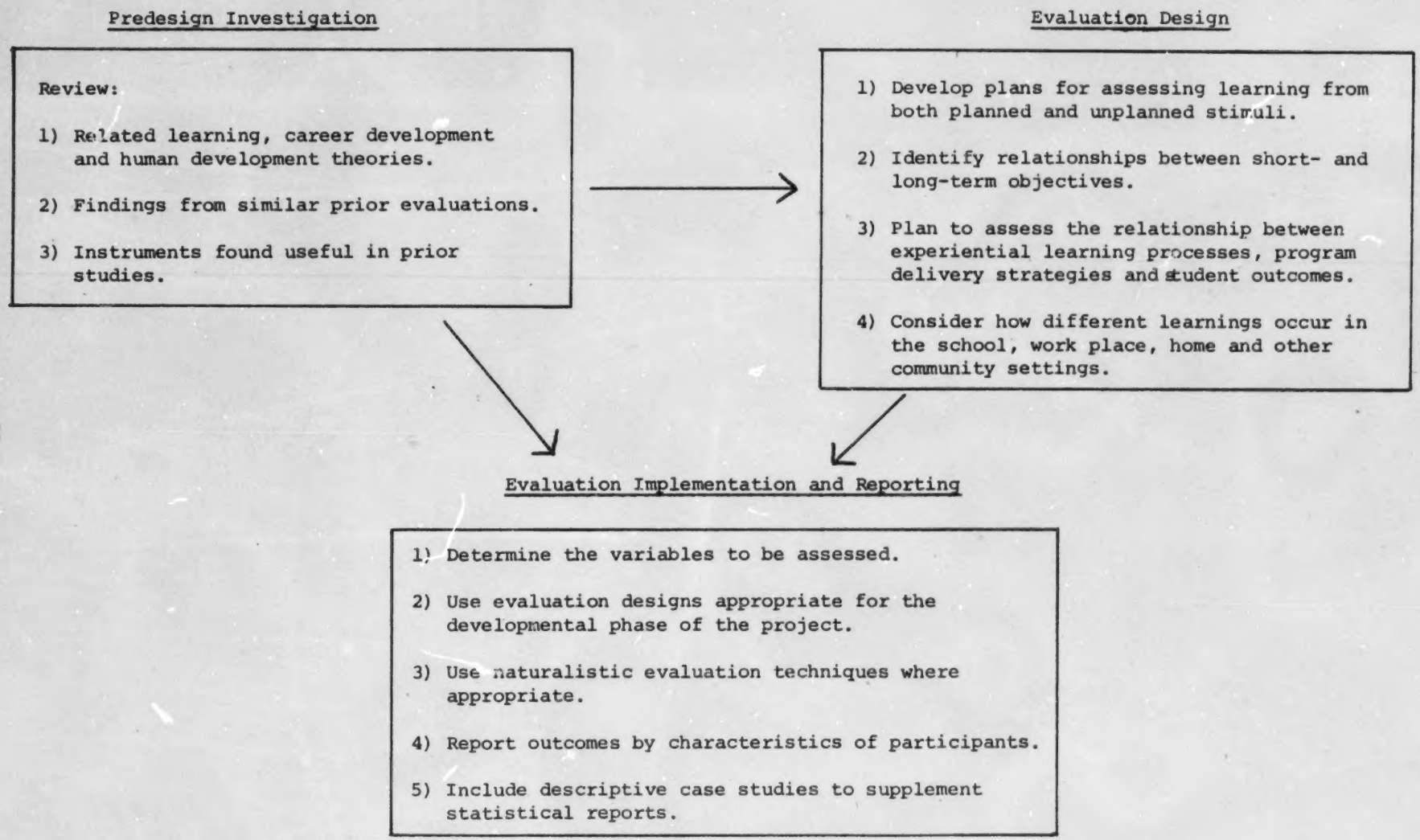
Proposed Framework for Assessing Experiential Learning Programs

It is tempting to identify problems involved in assessing experiential learning programs, such as those cited in the first section of this paper, and then to conclude with pious platitudes that improvements are surely needed. Such an approach may be prudent but not very helpful. Instead I would like to propose a preliminary framework for conducting assessments of experiential learning programs. Such a proposal is made with genuine humility since I, in no way, anticipate that these ideas will solve all the problems already cited. The approaches proposed here are not meant to be another evaluation model but rather a recommended framework for consideration by people already familiar with general evaluation principles. Key elements of this framework are depicted in Chart 1 and are discussed in the remainder of this section of the paper.

Predesign Investigations

Prior to developing an evaluation plan for an experiential learning program, it is useful to study factors both external and internal to the project. External factors include a review of some theories that may help organize and establish relationships among the separate objectives and student outcomes likely to be explored in the evaluation. Examples of such theories include social learning theory (Bandura, 1977), attribution theory (Bar-Tal, 1978) and motivation theory (Atkinson, 1964). A second area worth pursuing is that of evaluation findings resulting from prior programs similar to the one to be evaluated. In this respect it is equally important to determine which outcomes have been found not affected by such programs as to find which have shown

CHART 1: FRAMEWORK FOR EVALUATING EXPERIENTIAL LEARNING PROGRAMS


Predesign Investigation
Review:
1) Related learning, career development and human development theories.
2) Findings from similar prior evaluations.
3) Instruments found useful in prior studies.
Evaluation Design
1) Develop plans for assessing learning from both planned and unplanned stimuli.
2) Identify relationships between short- and long-term objectives.
3) Plan to assess the relationship between experiential learning processes, program delivery strategies and student outcomes.
4) Consider how different learnings occur in the school, work place, home and other community settings.
Evaluation Implementation and Reporting
1) Determine the variables to be assessed.
2) Use evaluation designs appropriate for the developmental phase of the project.
3) Use naturalistic evaluation techniques where appropriate.
4) Report outcomes by characteristics of participants.
5) Include descriptive case studies to supplement statistical reports.

significant changes. The third type of information useful before preparing an evaluation plan is the names and characteristics of evaluation techniques or instruments that have been used to assess outcomes similar to those expected in the program to be evaluated. As with program objectives, it is important to know which evaluation instruments revealed significant student gains and which ones did not. In the recent evaluation synthesis of K-12 career education programs sponsored by USOE Office of Career Education during 1975-76, Deborah Bonnet (1977, p.60) identified, for example, a commercially available instrument of career maturity that has revealed no significant growth in 32 out of 34 evaluation studies. Nevertheless, it continues to be widely used.

Designing the Evaluation

It has become fashionable for evaluation critics to attack goal statements and behavioral outcomes of many educational programs for lack of precision and coherence. Freeberg (1976), in writing about youth work training programs, has stated that "Instead of verifiable procedures, the prescriptions provided for specifying training objectives and associated measures appear to begin and end with purely consensual or rational decisions, usually leading to a laundry list of recommended performance outcomes that somehow appeal to the particular writer" (p. 537). In contrast, Freeberg borrows from criterion theory to recommend: 1) multiple behavioral indices, especially in the initial phases of criterion development; 2) specification of criteria along a temporal continuum to include short-term, intermediate and long-term measures, and 3) concurrent and predictive validity (the relevance of a criterion to others obtained at the same time and in the future).

In evaluating experiential learning programs where much of the learning takes place in an individualized manner at a community setting, it appears important to not only assess predesigned criteria, developed in the manner proposed by Freeberg, but also to be alert to assessing learning resulting from unplanned events which are experienced by the learner and reflected upon by the student and learning coordinator or teacher. An example of an unplanned stimulus for learning can occur in a situation where a student shows up late several days at a job site he or she was exploring and gets told by the employer not to return. An open discussion on this experience by a student and learning coordinator can produce new insights to help the young person modify both attitudes and future behavior.

Traditionally, employment and training programs have assessed whether a student has gained job entry skills, gained employment in the particular occupation and the salary received. More recently, researchers have stressed the need for also assessing occupational adaptability and transferable skills (Pratzner, 1978) that facilitate a person's ability to function successfully across a variety of careers and areas of adult life. Those involved in career education are faced with broader outcomes that include improved self-concept, work values, decision-making skills, use of leisure time and awareness of means available for continuing and recurrent education (Hoyt, 1978).

Service learning programs have focused on helping students to learn concern for others, ability to work with others, realistic attitudes toward different types of people and self-motivation to learn, participate and achieve (Hedin, 1979). Clearly the list is unending and

thus a critical element of the evaluator's role is to determine which objectives are central to the program under investigation and which have at least a fair chance of being achieved.

While it is important to look at learner outcomes, it is equally important to describe the delivery strategies being used to help students achieve these objectives. In the past, the delivery strategies have been generally ignored or described in broad terms such as career explorations and career counseling. In addition to depicting these strategies in a more concrete way, it would be useful to assess the experiential learning processes involved. In a recent research study based on a social learning theory framework, Owens and Owen (1979) identified a number of characteristics of experiential learning that high school students in eight Experience Based Career Education projects identified as important reasons why a particular community experience was an excellent or poor learning experience. Characteristics associated most with excellent learning experiences were, in descending order of importance:

1. Trying out tasks.
2. Being given real responsibility.
3. Listening and talking with people at community sites.
4. Receiving clear directions to follow.
5. Engaging in challenging tasks
6. Applying learning to new things.
7. Adult encouragement for doing tasks well.
8. Learning technical words and language associated with a career of interest.
9. Freedom to explore areas not planned in advance.

10. Personal interest of a supervisor.
11. Encouragement from family or friends.
12. Knowing ahead of time what would be expected.
13. Discussions of the student's experiences with staff and other students.

Characteristics students associated most with particular community experiences resulting in little or no learning were, in descending order of importance:

1. Boring tasks.
2. No personal opportunity to try out the work.
3. Too much repetition of an activity.
4. No opportunity to explore other areas of interest.
5. No opportunity to apply learning to new things.
6. Not knowing what would be expected.
7. No recognition from other workers for doing tasks well.
8. No opportunity to observe or talk to workers at the site.
9. Lack of clear directions to follow.
10. No opportunities to discuss experiences with other students.
11. Being too closely supervised.

Perceived importance of factors related to positive or negative learning experiences was not affected by student's sex. However, a significantly greater proportion of young females than males identified a better understanding of others as a learning outcome from their EBCE community site experiences. Students of both sexes felt that a major outcome of their community experiences was a better understanding of a specific career of interest.

More detailed research on some of these characteristics is likely to reveal important factors that motivate a young person to learn and to lead to recommended program improvements. As evaluators become more experienced in assessing special features of community based experiential learning programs, as compared to school based programs, they will also give attention to the diverse environments for learning that include the school, work place, home and other community settings (such as participation in a Boy Scouts of America program). The environments become critical to study in some cases because learning in one setting may reinforce or contradict learning in other settings. An example where confusion occurs is found when a young woman learns from her parents that her place as a future mother is in the home while at school she is encouraged to explore nontraditional careers.

At the expense of adding to the frustrations of an already overwhelmed reader wondering how he or she could possibly consider all of these aspects of student outcome evaluation, we feel it is useful to remind evaluators that, for some projects, it is equally important to assess program effects on the staff and on participating community resource persons. For example, it has been found, over several years, that participating employers see as benefits for their participation in Experience Based Career Education not only a chance to help students but, also, an opportunity to informally screen potential employees, to become better acquainted with the thinking of young people and gain prestige as a result of explaining their jobs to interested young people (Owens, Haenn and Fehrenbacher 1975).

Evaluation Implementation and Reporting

After engaging in the activities suggested above as part of the evaluation planning process it is necessary to touch base with reality in recognizing that the quickest way to an ineffective evaluation is to attempt to evaluate everything. Criteria are important for determining what variables to select for an intensive study, which may be explored in a more casual fashion and which go beyond the resources available. In selecting variables to be evaluated, the evaluator needs to consider factors such as: 1) legislative or administrative mandates related to the project; 2) the project's objectives; 3) areas likely to affect program modifications; 4) prior findings from related evaluations; 5) variables related to a theoretical explanation of why a program works; 6) the availability of valid measures to adequately assess potential variables; and 7) variables important for policy studies.

Persons reviewing evaluations of high school experiential learning programs have often commented on the inadequate research designs employed. Usually these criticisms reduce to a plea to employ an experimental research design such as that involving pre- and post-testing of an experimental and control group. The pros and cons involving the appropriateness of such designs have been thoroughly discussed elsewhere (Borouch and Rindskopf 1977, Guba 1978). The point we wish to make here is that decisions regarding appropriate designs need to be considered in light of the developmental cycle of a project and the purposes for the evaluation. In the case of most new programs, at least one year is needed to try out various delivery strategies, debug them and establish a stable program. Often a descriptive case study design is most

appropriate at this time. A second program phase may be needed to test the ability of the program to achieve its promised outcomes. In this phase an experimental or quasi-experimental design (Campbell and Stanley, 1973) or naturalistic approach (Guba, 1978) may be useful. Some programs progress to a third phase in which the project's delivery strategies have stabilized and the interest shifts to considerations about expanding the program or disseminating it to other districts. At this time people are usually interested in the longer-term impact of the program on former students and in issues related to the transportability of the project (i.e. program costs, staff training required and ways of selling the concept to other educators and the community). These concerns sometimes lead to graduate follow-up studies and cost-effectiveness studies.

Programs based on experiential learning attempt to place young people in "real world" environments in which they learn by directly engaging in activities and then reflecting on the meaning of their experiences. Such approaches are particularly appropriate for helping young people gain important competencies needed to be successful adults. Ideally, an effective evaluation of an experiential program would determine: 1) what learning took place; 2) how the learning occurred; 3) the different meanings that individuals attach to their experiences; and 4) the ability learners have gained to apply their new attitudes, knowledge and skills to new situations. Unfortunately, the author has not discovered an evaluation report that addresses all four areas.

Since experiential learning, as we have defined it, occurs in a real-world environment rather than in the artificial environment often found

in a classroom, it might be expected that naturalistic techniques would primarily be used to assess student performance. Unfortunately, most of the evaluation we reviewed relied on paper-and-pencil instruments to assess learner outcomes. Applied performance tests, designed to measure performance in a naturalistic setting, have not yet gained wide usage. Examples where such applied performance tests would be useful are in completing a job application, typing a business letter, demonstrating current life saving treatments in a first aid class, participating in a job interview simulation and in applying reading and mathematics to everyday work problems.

Examples of applied performance tests evolving from an experiential education program can be found in the Experience Based Career Education Program (EBCE) developed through the Northwest Regional Educational Laboratory. As a part of this EBCE program students master "survival skills" identified by the local community as essential for successful adult life in that community. At the EBCE demonstration site in Tigard, Oregon, for example, the community representatives identified the following 13 competencies: transacting business on a credit basis; maintaining a checking account in good order; providing adequate insurance for self, family and possessions; filing state and federal income tax; budgeting time and money effectively; maintaining good physical health and making appropriate use of leisure time; responding appropriately to fire, police and physical health emergencies; participating in the electoral process; understanding the basic structure and function of local, state and federal government; explaining one's legal rights and responsibilities; making appropriate use of public

agencies; making application for employment and successfully holding a job; and operating and maintaining an automobile. In most of these competencies, procedures have been identified for relevant volunteers of the community to certify successful completion of the competency. For example, a bank official certifies the checking account competency while an internal revenue service agent certifies the tax competency employing actual procedures and forms used by citizens in the community.

In addition to applied performance measures, a number of other nontesting evaluation techniques have been used by educators interested in evaluating student or program outcomes of experiential learning programs. These techniques include analysis of anecdotal records, checklists, content analysis of documents, drawings, interviews, journals, observations, peer nominations, photography, physical traces, questionnaires, rating scales, role-play, self-inventories, simulations and tape and videotape recordings. Although each of these techniques has its limitations, increased confidence in findings results when two or more evaluation techniques used to measure the same dimension produce supporting evidence.

So far we have presented a framework for conceptualizing, designing and implementing evaluations of experiential learning programs. A final consideration we wish to discuss is that of reporting evaluation findings.

Evaluation reports of experiential learning programs have seldom discussed program findings in relation to the characteristics of the learners involved. Seldom is it shown whether program delivery strategies have a differential effect depending on the learner's sex,

race, age, socioeconomic status or general achievement level. Likewise it would be useful, when possible, to relate experiential learning characteristics to program delivery strategies and student outcomes. For example, positive reinforcement by community resource people encountered in career explorations may be a key influence on some youth in deciding which particular career field to enter. Such findings can help us go beyond learning that a student has made a career choice, to understanding predominant influences that led to that learning.

Statistical reports seldom communicate the flavor of what experiential elements or delivery systems were particularly effective or ineffective in contributing to one or more learner outcomes. Carefully planned and documented student case study procedures have been developed (Fehrenbacher, Owens and Haenn, in process) that complement statistical findings in conveying to staff, students and community people the individualized treatment that should be part of an experiential learning program.

Recent and Emerging Developments

At this point you may feel that the framework proposed above falls far short of solving many of the problems discussed in the first section of this paper. I would agree with you! Fortunately, there are some recent or emerging developments giving hope for the future that may be of interest to you. These developments include those involved in: 1) the dissemination of information about effective experiential learning programs; 2) research and evaluation on experiential learning; and 3) new assessment techniques.

Program Dissemination

In 1967 the National Commission on Resources for Youth (Kohler and Dollar, 1976) was formed in New York to encourage development and national dissemination of youth participation programs. This commission has helped many school districts interested in developing programs in which young people have direct involvement in planning and operating programs that provide such personal services as tutoring of younger students or community improvement projects. The Commission maintains and disseminates descriptions about many such programs. A similar clearinghouse of information was established in 1978 at the Northwest Regional Educational Laboratory (NWREL) to serve the Pacific Northwest. This clearinghouse, called the Experiential Learning Information Exchange, will provide information about programs, resource people and research related to experiential learning.

In 1978 a Collaborative for Experiential Learning was established by Murray Durst to serve as a link for persons interested in experiential education and has begun teacher assistance centers at six colleges or universities around the country. These centers help teachers develop and use experiential learning approaches with their classes.

Research and Evaluation

The Education and Work Program of NWREL and staff at the National Center for Research in Vocational Education at the Ohio State University are engaged in research on essential characteristics associated with experiential learning. Both agencies have developed a review of the literature on experiential learning (D. Coleman, 1978 and Druian, Owen and Owens, in press). The NWREL study also involves a description of common elements and processes of experiential learning determined from

interviews with staff of five or six nationally recognized programs such as work experience programs, Foxfire, Experience Based Career Education, Outward Bound and the Explorer Scouts Program of the Boy Scouts of America.

Last year a national evaluation of approximately 30 secondary public and private experiential learning programs was initiated at the Center for Research on Youth Development at the University of Minnesota (Conrad and Hedin, 1979). This national evaluation is using student survey instruments and project c se studies to determine the setting, learning processes and outcomes resulting from these experiential learning programs.

A Research on Evaluation project at NWREL that began last year is designed to develop new evaluation methodology. The project is currently examining other disciplines such as architecture, law, geography, investigative journalism, operations research, philosophy and poetic criticism as metaphors for educational evaluation. Some of these disciplines use techniques that may have important payoffs for the assessment of experiential learning programs.

Assessment Techniques

In addition to the work of the Clearinghouse for Applied Performance Testing at NWREL in developing and disseminating information on assessment techniques, the National Center for Research in Vocational Education has developed three handbooks on assessment under funding provided by the National Institute of Education. One is intended to assist practitioners to develop local instruments, another deals with the use of unobtrusive measures in program evaluation and the third is a

description of recent assessment instruments developed by researchers around the country that would be useful in experiential learning programs. These three handbooks are due for release soon.

The problems and challenges faced in effectively assessing competencies involved in experiential learning programs are enormous. The state of the art indicates to me that we have much to learn about such assessment. With the work and support of you and others involved in nontraditional testing techniques I feel that we can make a significant contribution to public and private education in the years ahead.

BIBLIOGRAPHY

Atkinson, J. W. An introduction to motivation. Princeton, N.J.: Van Nostrand, 1964.

Bandura, A. Social learning theory. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Bar-Tal, D. Attributional analysis of achievement-related behavior. Review of Educational Research, 1978, 48(2), 259-271.

Bonnet, D. G. A synthesis of results and programmatic recommendations emerging from career education evaluations in 1975-76. (Office of Education, Office of Career Education). Washington, D.C.: U.S. Government Printing Office, 1977.

Borouch, R. F. & Rindskopf, D. On randomized experiments, approximations to experiments and data analysis. In L. Rutman (Ed.) Evaluation and research methods: A research guide. Beverly Hills, CA: Sage, 1977.

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963.

Coleman, D. (Ed.) Experiential education linking education and work: Annotated bibliography. (Review Copy) Columbus, OH: National Center for Research in Vocational Education, 1978.

Coleman, J. S. Differences between experiential and classroom learning. In Keeton, M. T. (Ed.), Experiential Learning: Rationale, Characteristics and Assessment. San Francisco: Jossey-Bass, 1976.

Conrad, D. & Hedin, D. Testing the claims for service-learning. Synergist, 1979, 7(3), 28-30.

Druian, G., Owen, S. K., & Owens, T. R. Review of selected literature pertaining to experiential education. Portland, OR: Northwest Regional Educational Laboratory, (in press).

Fehrenbacher, H. L., Owens, T. R. & Haenn, J. F. The use of student case study methodology in program evaluation. Journal of Research and Development in Education (in press).

Freeberg, N. E. Criterion measures for youth-work training programs: The development of relevant performance dimensions. Journal of Applied Psychology, 19176, 61(5), 537-545.

Guba, E. G. Toward a methodology of naturalistic inquiry in educational evaluation. CSE Monograph Series in Evaluation, 1178, (8). Los Angeles: University of California, Center for the Study of Evaluation.

Hedin, D. P. Teenage health educators: An action learning program to promote psychological development. Unpublished doctoral dissertation, University of Minnesota, 1979.

Hoyt, K. B. The concept of collaboration in career education. Washington, D.C.: U.S. Government Printing Office, 1978.

Kohler, M. C. & Dollar, B. An antidote to alienation. The Center Magazine. Center for the Study of Democratic Institutions, 9, 1976, 20-27.

Miguel, R. C. (Ed.) Experiential Education Guidelines. Developed by Experiential Education Advisory Panel. Columbus, OH: The National Center for Research in Vocational Education, 1978.

Owens, T., Haenn, J. & Febrenbacher, H. FY 75 Final evaluation report of the NWREL Experience-Based Career Education program. Portland, OR: Northwest Regional Educational Laboratory, 1975.

Owens, T. R. & Owen, S. K. Enhancing the quality of community learning experiences. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April, 1979.

Owens, T. R., Owen, S. K., & Druian, G. Experiential learning programs: Synthesis of findings and proposed framework for future evaluations. Portland, OR: Northwest Regional Educational Laboratory, 1979.

Pratzner, F. C. Occupational adaptability and transferable skills. Columbus, OH: The National Center for Research in Vocational Education, 1978.